COPY OF PAPERS ORIGINALLY FILED

PATENT

Docket No. 1232-4563

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants	:	Yasuo Okutani and Masayuki Yamada	Group Art Unit: 2641
Serial No.	:	09/386,052	Examiner: Angela A. Armstrong
Filed	:	August 30, 1999	
For	:	**SPEECH SYNTHESIZING APPARATUS AND METHOD, AND STORAGE MEDIUM THEREOF**	

TRANSMITTAL OF CERTIFIED TRANSLATION OF PRIORITY APPLICATION

COMMISSIONER OF PATENTS
Washington, D.C. 20231

Sir:

A Preliminary Amendment was submitted on May 14, 2002 in the above identified application in which Applicants pointed a reference of record- Manwaring (USP 6,188,984)- is not prior art against this application because Manwaring was filed on November 17, 1998. While the present application was filed on August 30, 1999, priority is claimed to Japanese application JP 10-245951, filed on August 31, 1998.

In this regard, submitted herewith is a copy of the certified English translation of Japanese Application No. 10-245951.

In view of the foregoing, and as set forth in further detail the Preliminary Amendment, Applicants respectfully submit that the application is in condition for allowance which action is earnestly requested.

In the event that a telephone conference would facilitate prosecution of the application in any way, the Examiner is invited to contact the undersigned at the number provided. An early and favorable examination on the merits is respectfully requested.

COPIES
ORIGINALLY FILED

PATENT **Docket No. 1232-4563**

AUTHORIZATION

No additional fees are believe due at this time, however the Commissioner is hereby authorized to charge any additional fees which may be required for this paper, or credit any overpayment, to Deposit Account No. 13-4500, Order No. 1232- 4563.

Respectfully submitted,

MORGAN & FINNEGAN, L.L.P.

By: _____
Stephen J. Manetta
Reg. No. 40,426

Date: May 28, 2002

Mailing Address:

MORGAN & FINNEGAN, L.L.P.
345 Park Avenue
New York, New York 10154
(212) 758-4800 Telephone
(212) 758-6849 Facsimile

704693 v1

2641

Docket No. 1232-4563

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): Yasuo Okutani and Masayuki Yamada

Group Art Unit: 2641

Serial No.: 09/386,052

Examiner: Angela A. Armstrong

Filed: August 30, 1999

For: SPEECH SYNTHESIZING APPARATUS AND METHOD, AND STORAGE MEDIUM THEREOF

CERTIFICATE OF MAILING (37 C.F.R. §1.8(a))

Commissioner for Patents
Washington, DC 20231

Sir:

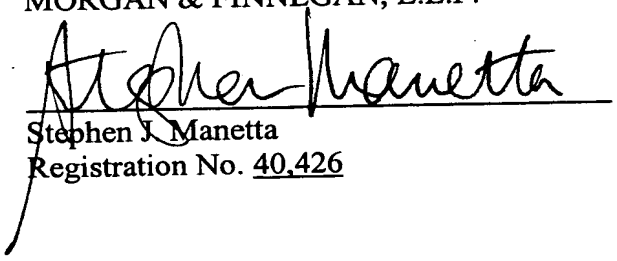
I hereby certify that the attached:

1. Transmittal Of Certified Translation Of Priority Application;
2. Certified English translation of Japanese Application No. 10-245951, and
3. Return Receipt Postcard.

along with any paper(s) referred to as being attached or enclosed and this Certificate of Mailing are being deposited with the United States Postal Service on date shown below with sufficient postage as first-class mail in an envelope addressed to the: Commissioner for Patents, Washington, DC 20231.

Respectfully submitted,
MORGAN & FINNEGAN, L.L.P.

Dated: May 28, 2002

By: 
Stephen J. Manetta
Registration No. 40,426

Correspondence Address:
MORGAN & FINNEGAN, L.L.P.
345 Park Avenue
New York, NY 10154-0053
(212) 758-4800 Telephone
(212) 751-6849 Facsimile

10-245951

[Type of the Document] SPECIFICATION

[Title of the Invention]

SPEECH SYNTHESIZING APPARATUS AND METHOD, AND STORAGE MEDIUM

[Claims]

[Claim 1] A speech synthesizing apparatus for synthesizing a speech waveform, characterized by having:

storage means for storing vocalization data in which attribute information has been assigned to phoneme data;

retrieval means for retrieving phoneme data, in accordance with predetermined retrieval conditions, from the vocalization data stored in said storage means;

assigning means for assigning a penalty, based upon at least some of said attribute information, in a set of the phoneme data retrieved by said retrieval means;

selection means for selecting, from the set of said phoneme data, and based upon the penalty assigned by said assigning means, phoneme data to be employed in synthesis of a speech waveform.

[Claim 2] A speech synthesizing apparatus according to claim 1, characterized in that said attribute information includes phoneme label, phoneme boundary, fundamental frequency, power and phoneme duration information.

[Claim 3] A speech synthesizing apparatus according to claim 1 or 2, characterized in that said retrieval

means retrieves phoneme data that satisfies a specified phoneme environment.

[Claim 4] A speech synthesizing apparatus according to claim 2, characterized in that said retrieval means retrieves phoneme data that satisfies a specified phoneme environment and fundamental frequency.

[Claim 5] A speech synthesizing apparatus according to claim 2, characterized in that said assigning means assigns a penalty in relation to power and phoneme duration of each item of phoneme data.

[Claim 6] A speech synthesizing apparatus according to claim 5, characterized in that said assigning means:

sorts the items of phoneme data in order of decreasing power and assigns a power-related penalty on the basis of the order obtained by sorting, in such a manner that a small penalty is assigned to phoneme data whose power is close to an average value; and

sorts the items of phoneme data in order of decreasing phoneme duration and assigns a phoneme-duration-related penalty on the basis of the order obtained by sorting, in such a manner that a small penalty is assigned to phoneme data whose phoneme duration is close to an average value.

[Claim 7] A speech synthesizing apparatus according to claim 1, characterized by further having:

alternate retrieval means for retrieving phoneme data that satisfies part of said retrieval environment

in a case where phoneme data that satisfies a specific environment in said retrieval means does not exist; and

counting means for counting the items of phoneme data for every phoneme environment of phoneme data retrieved by said alternate retrieval means;

wherein said assigning means assigns a penalty, on the basis of at least some of said attribute information, to the phoneme data retrieved by said alternate retrieval means, and assigns a penalty based upon a numerical value obtained by said counting means.

[Claim 8] A speech synthesizing method for storing, in storage means, vocalization data in which attribute information has been assigned to phoneme data, and synthesizing a speech waveform using the phoneme data stored in said storage means, characterized by having:

a retrieval step of retrieving phoneme data, in accordance with predetermined retrieval conditions, from vocalization data stored in said storage means;

an assigning step of assigning a penalty, based upon at least some of said attribute information, in a set of the phoneme data retrieved at said retrieval step; and

a selection step of selecting, from the set of said phoneme data, and based upon the penalty assigned at said assigning step, phoneme data to be employed in synthesis of a speech waveform.

[Claim 9] A speech synthesizing method according to

claim 8, characterized in that said attribute information includes phoneme label, phoneme boundary, fundamental frequency, power and phoneme duration information.

[Claim 10] A speech synthesizing method according to claim 8 or 9, characterized in that said retrieval step retrieves phoneme data that satisfies a specified phoneme environment.

[Claim 11] A speech synthesizing method according to claim 9, characterized in that said retrieval step retrieves phoneme data that satisfies a specified phoneme environment and fundamental frequency.

[Claim 12] A speech synthesizing method according to claim 9, characterized in that said assigning step assigns a penalty in relation to power and phoneme duration of each item of phoneme data.

[Claim 13] A speech synthesizing method according to claim 12, characterized in that said assigning step:

sorts the items of phoneme data in order of decreasing power and assigns a power-related penalty on the basis of the order obtained by sorting, in such a manner that a small penalty is assigned to phoneme data whose power is close to an average value; and

sorts the items of phoneme data in order of decreasing phoneme duration and assigns a phoneme-duration-related penalty on the basis of the order obtained by sorting, in such a manner that a small

penalty is assigned to phoneme data whose phoneme duration is close to an average value.

[Claim 14] A speech synthesizing method according to claim 8, characterized by further having:

an alternate retrieval step of retrieving phoneme data that satisfies part of said retrieval environment in a case where phoneme data that satisfies a specific environment at said retrieval step does not exist; and

a counting step of counting the items of phoneme data for every phoneme environment of phoneme data retrieved at said alternate retrieval step;

wherein said assigning step assigns a penalty, on the basis of at least some of said attribute information, to the phoneme data retrieved at said alternate retrieval step, and assigns a penalty based upon a numerical value obtained at said counting step.

[Claim 15] A storage medium storing a control program for causing a computer to synthesize a speech waveform using phoneme data, to which attribute information has been assigned, stored in storage means, characterized by having:

code of a retrieval step of retrieving phoneme data, in accordance with predetermined retrieval conditions, from vocalization data stored in said storage means;

code of an assigning step of assigning a penalty, based upon at least some of said attribute information,

in a set of the phoneme data retrieved at said retrieval step; and

code of a selection step of selecting, from the set of said phoneme data, and based upon the penalty assigned at said assigning step, phoneme data to be employed in synthesis of a speech waveform.

[Claim 16] A storage medium according to claim 15, characterized by further having:

code of an alternate retrieval step of retrieving phoneme data that satisfies part of said retrieval environment in a case where phoneme data that satisfies a specific environment at said retrieval step does not exist; and

code of a counting step of counting the items of phoneme data for every phoneme environment of phoneme data retrieved at said alternate retrieval step;

wherein the code of said assigning step includes code of processing for assigning a penalty, on the basis of at least some of said attribute information, to the phoneme data retrieved at said alternate retrieval step, and assigns a penalty based upon a numerical value obtained at said counting step.

[Detailed Description of the Invention]

[0001]

[Technical Field to which the Invention Belongs]

This invention relates to an speech synthesizing apparatus having a database for managing phoneme data,

in which the apparatus performs speech synthesis using the phoneme data managed by the data, to a method of synthesizing speech, and to a storage medium storing a program for implementing this method.

[0002]

[Prior Art]

A method of speech synthesis by waveform editing (which will be referred to as the "waveform editing synthesis method" below) exists in the prior art as a method of synthesizing speech. With the waveform editing synthesis method, rhythm is changed by a pitch-synchronized waveform superposition method in which waveform segments of one to several pitches are joined together in conformity with a desired pitch spacing. An advantage of the waveform editing synthesis method is that the synthesized speech obtained is more natural than that provided by a synthesis method based upon parameters. On the other hand, a problem is that the allowable range for the change in rhythm is narrow.

[0003]

Accordingly, sound quality is improved by preparing speech data of a wide variety of variations, selecting these properly and using them. Information such as the phoneme environment (the phoneme that is the object of synthesis or several phonemes on both sides thereof) and the fundamental frequency F_0 is used as the criteria for selecting the speech data.

[0004]

[Problem to be solved by the Invention]

However, the conventional method of synthesizing speech described above involves a number of problems.

[0005]

By way of example, even if a database contains a plurality of items of phoneme data which satisfy a certain phoneme environment and the fundamental frequency F_0, the phoneme unit used in synthesis is one phoneme unit (e.g., the phoneme unit that appears in the database first) selected randomly from these items of phoneme data. Since the database is a set of speech uttered by human beings, all of the phoneme data is not necessarily stable (i.e., not necessarily of good quality). The database may contain phoneme data that is the result of mumbling, a halting voice, slowness of speech or hoarseness. If one item of phoneme data is selected carelessly from such a set of data, naturally there is the possibility that sound quality will decline when synthesized speech is generated.

[0006]

The present invention has been devised in view of the foregoing problems and its object is to provide a speech synthesizing apparatus and method capable of appropriately selecting phoneme data used in speech synthesis and of suppressing a decline in sound quality in speech synthesis, as well as a storage medium storing

a program for implementing this control method.

[0007]

[Means for Solving the Problem]

A speech synthesizing apparatus embodying the present invention for attaining the foregoing object has the following construction, by way of example:

A speech synthesizing apparatus for synthesizing a speech waveform, having:

storage means for storing vocalization data in which attribute information has been assigned to phoneme data;

retrieval means for retrieving phoneme data, in accordance with predetermined retrieval conditions, from the vocalization data stored in the storage means;

assigning means for assigning a penalty, based upon at least some of the attribute information, in a set of the phoneme data retrieved by the retrieval means;

selection means for selecting, from the set of the phoneme data, and based upon the penalty assigned by the assigning means, phoneme data to be employed in synthesis of a speech waveform.

[0008]

A speech synthesizing method according to another embodiment of the present invention for attaining the foregoing object has the following steps, by way of example:

A speech synthesizing method for storing, in

storage means, vocalization data in which attribute information has been assigned to phoneme data, and synthesizing a speech waveform using the phoneme data stored in said storage means, having:

a retrieval step of retrieving phoneme data, in accordance with predetermined retrieval conditions, from vocalization data stored in the storage means;

an assigning step of assigning a penalty, based upon at least some of the attribute information, in a set of the phoneme data retrieved at the retrieval step; and

a selection step of selecting, from the set of the phoneme data, and based upon the penalty assigned at the assigning step, phoneme data to be employed in synthesis of a speech waveform.

[0009]

Further, in accordance with the present invention, there is provided a storage medium storing a control program for causing a computer to implement the above-described speech synthesizing method.

[0010]

[Embodiment of the Invention]

Preferred embodiments of the present invention will now be described with reference to the accompanying drawings.

[0011]

[First Embodiment]

Fig. 1 is a block diagram illustrating the construction of a speech synthesizing apparatus according to a first embodiment of the present invention. In Fig. 1, numeral 101 denotes a control memory (ROM) 101 which stores a control program for causing a computer to implement control in accordance with a control procedure shown in the flowchart of Fig. 3. Numeral 102 denotes a central processing unit for executing processing such as decisions and calculations in accordance with the control procedure retained in the control memory 101. Numeral 103 denotes a memory (RAM) which provides a work area for when the central processing unit 102 executes various control operations. Allocated to the memory 103 are an area 202 for holding the results of phoneme retrieval, an area 204 for holding the results of penalty assignment, an area 207 for holding the results of sorting, and an area 209 for holding representative phoneme data. These areas will be described with reference to Fig. 2. Numeral 104 denotes a disk device which, in this embodiment, uses a hard disk. The disk device 104 stores a database 200 described later with reference to Fig. 2. The data of database 200 is stored in memory 103 when the data is used. Numeral 105 denotes a bus which connects the components mentioned above.

[0012]

The speech synthesizing apparatus of this

embodiment uses information such as the phoneme environment and fundamental frequency to select the appropriate phoneme data from speech data that has been recorded in the database 200 and performs waveform editing synthesis employing the selected data. Below will be described a case where selection of phoneme data is carried out using the phoneme environment (the phoneme of interest and one phoneme on each side thereof, these being referred to as a so-called "triphone") and information concerning the average fundamental frequency of the phoneme as criteria for selecting phoneme data.

[0013]

Fig. 2 is a block diagram illustrating functional arrangements of a speech synthesizing apparatus according to this embodiment, which relates to phoneme data selection processing for selecting the optimum phoneme data from a set of phoneme data in which the phoneme environments and fundamental frequencies are identical.

[0014]

In this Figure, numeral 200 denotes a database which stores vocalization data in which a phoneme label, phoneme boundary and fundamental frequency, power and phoneme duration are have been assigned to each item of phoneme data. Numeral 201 denotes a phoneme retrieval unit for retrieving phoneme data, which satisfies a

specific phoneme environment and fundamental frequency, from the database 200. Numeral 202 denotes a retrieved-result storage area for storing a set of phoneme data, namely the results of retrieval performed by the phoneme retrieval unit 201. Numeral 203 denotes a power-penalty assignment processing unit which assigns a penalty related to power to each item of phoneme data of the set of phoneme data stored in the retrieved-result storage area 202. Numeral 204 denotes an area which holds the results of the assignment of penalties to the phoneme data. Numeral 205 denotes a duration-penalty assignment processing unit which assigns a penalty relating to phoneme duration to each items of phoneme data.

[0015]

Numeral 206 denotes a sorting processing unit which subjects the set of phoneme data to sorting processing regarding specific information (power or phoneme duration, etc.) when a penalty is assigned. Numeral 207 denotes an area which holds the results of sorting. In regard to the results obtained by assigning penalties, numeral 208 denotes a data determination processing unit which selects phoneme data having the smallest penalty as representative phoneme data. Numeral 209 denotes an area which holds the representative phoneme data that has been decided.

[0016]

From the speech synthesizing processing set forth

above, processing for selecting phoneme data implemented by the above-described functional arrangement will be discussed next. Fig. 3 is a flowchart illustrating a procedure relating to phoneme data selection processing for selecting the optimum phoneme data from the set of phoneme data having identical phoneme environments and fundamental frequencies.

[0017]

First, at step S301, all phoneme data that satisfies the specified phoneme environment and fundamental frequency is extracted from the database 200 and the set of phoneme data retrieved is stored in the retrieved-result storage area 202. Next, at step S302, the power-penalty assignment processing unit 203 assigns power-related penalties to the set of phoneme data that has been stored in retrieved-result storage area 202.

[0018]

The guideline involving penalties is to assign penalties to phoneme data having power values that depart from an average value of power because the goal is to select phoneme data having an average value of power within the set of phoneme data. The power-penalty assignment processing unit 203, therefore, instructs the sorting processing unit 206 to sort the phoneme data set, which has been extracted from the retrieved-result storage area 202 that holds the results of retrieval, based upon values of power. Power referred to here may

be the power of the phoneme data or the average power per unit of time.

[0019]

The sorting processing unit 206 responds by sorting the phoneme data set based upon power and storing the results in the sorting-result storage area 207. The power-penalty assignment processing unit 203 waits for sorting to end and then assigns a penalty to the sorted phoneme data that has been stored in area 207. A penalty is assigned in accordance with the guideline mentioned above. For example, among items of phoneme data that have been sorted in order of decreasing power, a penalty (e.g., 2.0 points) is added onto phoneme data whose power values fall within the smaller one-third of values and onto phoneme data whose power values fall within the larger one-third of values. In other words, a penalty is assigned to phoneme data other than the middle one-third of phoneme data. The result of assigning the penalty is stored in the area 204 which holds the results of the assignment of penalties. Control then proceeds to step S303.

[0020]

Next, at step S303, the phoneme duration-penalty assignment processing unit 205 assigns a penalty relating to phoneme duration through a procedure similar to that of the power-penalty assignment processing unit. Specifically, the phoneme duration-penalty assignment

processing unit 205 instructs the sorting processing unit 206 to perform sorting based upon phoneme duration and stores the results in sorting-result storage area 207. On the basis of the sorted results, the phoneme duration-penalty assignment processing unit 205 adds a penalty (e.g., 2.0 points) onto phoneme data whose phoneme durations fall within the smaller one-third of durations and onto phoneme data whose phoneme durations fall within the larger one-third of durations. The results obtained by the assignment of the penalty are retained in area 204 which holds the results of the assignment of penalties. Control then proceeds to step S304.

[0021]

In step S304, the data determination processing unit 208 determines a representative phoneme unit in terms of the phoneme environment and fundamental frequency currently of interest. Here the results of penalty assignment based upon power and phoneme duration and stored in area 204 are delivered to the sorting processing unit 206 and the sorting processing unit 206 is instructed to sort the results by penalty value. The sorting processing unit 206 performs sorting on the basis of the two types of penalties relating to power and phoneme duration and stores the sorted results in sorting-result storage area 207. When sorting processing ends, the data determination processing unit

208 selects phoneme data having the smallest penalty and stores it in area 209, which is for storing representative phoneme data, for the purpose of employing this data as representative phoneme data. If a plurality of phoneme units having the minimum penalty value appear, the data determination processing unit 208 selects the phoneme unit located at the head of the sorted results. This is equivalent to selecting one appropriate phoneme unit from those having the smallest penalty.

[0022]

Thus, in accordance with the first embodiment, the optimum phoneme data is selected, based upon a penalty relating to power and a penalty relating to phoneme duration, from a phoneme data set in which the phoneme environments and fundamental frequencies are identical.

[0023]

[Second Embodiment]

The first embodiment has been described in regard to a case where the phoneme environment (the "triphone", namely the phoneme of interest and one phoneme on each side thereof) and the average fundamental frequency F_0 of the phoneme are used as criteria for selecting phoneme data. However, in instances where the triphone of a combination not contained in the database is required, the need arises to use an alternate "left-phone" (a phoneme environment comprising the phoneme of

interest and the phoneme to its left), "right-phone" (a phoneme environment comprising the phoneme of interest and the phoneme to its right) or "phone" (the phoneme of interest alone). In the second embodiment, therefore, there will be described a case where selection of phoneme data other than a specified triphone (such selected phoneme data will be referred to as a "triphone substitute") is taken into account.

[0024]

Fig. 4 is a block diagram illustrating functional arrangements of a speech synthesizing apparatus according to the second embodiment, which relates to phoneme data selection processing for selecting the optimum phoneme data from a set of phoneme data in which the phoneme environments and fundamental frequencies are identical. This embodiment differs from the first embodiment in Fig. 2 in that the apparatus is additionally provided with a processing unit 410 for assigning element-number penalty. Other areas or units 400 to 409 correspond to the areas or units 200 to 209, respectively, of Fig. 2. The processing unit 410 assigns a penalty in dependence upon the number of elements in a set of phoneme data.

[0025]

Described next will be a procedure, in speech synthesizing processing, relating to phoneme data selection processing, which is implemented by the above-

described functional blocks, for selecting optimum phoneme data from a set of phoneme data having identical phoneme environments and fundamental frequencies. Fig. 5 is a flowchart illustrating a procedure according to the second embodiment relating to phoneme data selection processing for selecting the optimum phoneme data from the set of phoneme data having identical phoneme environments and fundamental frequencies.

[0026]

Steps S501 to S503 are similar to steps S301 to S303 (Fig. 3) in the first embodiment. It should be noted that if a specified triphone does not exist in the database, the triphone retrieval at step S501 involves the retrieval of the alternate candidates left-phone, right-phone and phone (these are referred to as "triphone substitutes").

[0027]

In the second embodiment, use of a triphone substitute means that a specified triphone does not exist. As long as a specified triphone is contained in the database, however, this triphone is adopted. At step S504, therefore, it is determined whether only a triphone substitute has been obtained as the result of retrieval. If the specified triphone has been retrieved, control proceeds to step S506. When the specified triphone is retrieved, therefore, processing similar to that of the first embodiment is executed. If

it is determined at step S504 that only a triphone substitute has been retrieved, on the other hand, control proceeds to step S505. Here the processing unit 505 assigns a penalty in dependence upon the numbers of elements in the set of phoneme data. In a case where the specified triphone is absent, the processing unit 505 counts the numbers of elements contained in the phoneme data set, the count being performed per each triphone phoneme environment (an environment comprising the phoneme concerned and one phoneme on each side thereof) of the alternate candidate left-phone (or right-phone or phone). Here, if the number of items of phoneme data of an applicable triphone phoneme environment is small (two or less), then the processing unit 505 adds a penalty (0.5 points) onto all of the phoneme data concerned. In other words, the processing unit 505 judges that data having only a low frequency of appearance in a sufficiently large database is not reliable.

[0028]

For example, consider a case where a triphone t.A.k. does not exist in the database and is to be replaced by a left-phone t.A.*. If two triphones t.A.p. and 20 triphones t.A.t. exist in the database, allocating a triphone substitute, which is to replace the triphone t.A.k., from among triphones t.A.t. of which 20 exist will provided a higher probability of

obtaining phoneme data of good quality.

[0029]

If a penalty based upon number of elements is thus assigned, the result is stored in area 504, which is for holding the results of penalty assignment, and then control proceeds to step S506. Step S505 involves processing equivalent to that of step S304 in the first embodiment. In the second embodiment, a penalty based upon number of elements is assigned in addition to the penalty based upon power and the penalty based upon phoneme duration. As a result, phoneme data is selected upon taking all of these three penalties into consideration. In a case where a specific triphone is retrieved and processing proceeds directly from step S504 to step S506, penalty based upon number of elements is not taken into account.

[0030]

Thus, in accordance with the second embodiment, it is possible to select the proper phoneme data inclusive of triphones that can be alternates.

[0031]

In the embodiments set forth above, a case has been described in which penalty assignment processing is executed in order of power penalty and phoneme-duration penalty (and then element-number penalty). However, this does not impose a limitation upon the present invention, for the processing may be executed in any

order. Further, an arrangement may be adopted in which these penalty assignment processing operations are executed concurrently.

[0032]

Further, in each of the foregoing embodiments, 2.0 points is adopted as the penalty value for the power and phoneme-duration penalties. However, this does not impose a limitation upon the present invention, for it is obvious that a suitable value may be set. In addition, equal penalties need not be applied as the penalties relating to both characteristics.

[0033]

In the second embodiment, a case in which 0.5 is set as the value of the element-number penalty is described. However, this does not impose a limitation upon the present invention, for a suitable value may be set.

[0034]

Furthermore, in each of the foregoing embodiments, a case is described in which a penalty is assigned to the one-third of phoneme data starting from smaller values (or to the one-third of phoneme data starting from larger values) in regard to the sorted results. However, this does not impose a limitation upon the present invention. For example, it is possible to change the method of penalty assignment depending upon the number of items of phoneme data or the properties of

the phoneme data contained in the database. In such case a penalty may be assigned to data for which the difference relative to an average value is greater than a threshold value.

[0035]

Further, in the foregoing embodiments, there is described a method of selecting representative phoneme data in which the target is a phoneme data set that satisfies a specific phoneme environment and fundamental frequency. However, this does not impose a limitation upon the present invention. For example, it is possible to use a phoneme data set for which the matter of interest is solely the phoneme environment and to adopt the fundamental frequency as a factor for assigning a penalty.

[0036]

Further, in each of the above embodiments, there is described a method of selecting a representative phoneme unit on demand, wherein the target is a phoneme data set that satisfies a specific phoneme environment and fundamental frequency. However, an arrangement may be adopted in which a phoneme lexicon obtained by applying the processing of the first embodiment in advance is created based upon all conceivable phoneme environments and fundamental frequencies.

[0037]

Further, in each of the foregoing embodiments, a

case in which the sorting processing unit and the area for holding the sorted results are designed for general-purpose use. However, this does not impose a limitation upon the present invention. For example, an arrangement may be adopted in which there is provided a sorting processor exclusively for the processing unit that assigns the power penalties and a sorting processor exclusively for the processing unit that assigns the phoneme-duration penalties.

[0038]

In each of the foregoing embodiments, a case in which the areas for storing data are implemented by memory (RAM) is described. However, this does not impose a limitation upon the present invention because any storage media may be used.

[0039]

Further, in each of the foregoing embodiments, a case in which the components are constituted by the same computer is described. However, this does not impose a limitation upon the present invention because these components may be implemented by computers or processors distributed over a network.

[0040]

Further, in each of the foregoing embodiments, a case in which a program is stored in a control memory (ROM) is described. However, this does not impose a limitation upon the present invention because the

program may be stored in any storage media. The same operations performed by the program may be carried out by circuitry.

[0041]

The present invention can be applied to a system constituted by a plurality of devices or to an apparatus comprising a single device. Furthermore, it goes without saying that the invention is attained by supplying a storage medium storing the program codes of the software for performing the functions of the foregoing embodiments to a system or an apparatus, reading the program codes with a computer (e.g., a CPU or MPU) of the system or apparatus from the storage medium, and then executing the program codes.

[0042]

In this case, the program codes read from the storage medium implement the novel functions of the invention, and the storage medium storing the program codes constitutes the invention.

[0043]

Further, the storage medium, such as a floppy disk, hard disk, optical disk, magneto-optical disk, CD-ROM, CD-R, magnetic tape, non-volatile type memory card or ROM can be used to provide the program codes.

[0044]

Furthermore, besides the case where the aforesaid functions according to the embodiments are implemented

by executing the program codes read by a computer, it goes without saying that the present invention covers a case where an OS or the like running on the computer performs a part of or the entire process in accordance with the designation of program codes and implements the functions according to the embodiments.

[0045]

It goes without saying that the present invention further covers a case where, after the program codes read from the storage medium are written in a function expansion board inserted into the computer or in a memory provided in a function expansion unit connected to the computer, a CPU or the like contained in the function expansion board or function expansion unit performs a part of or the entire process in accordance with the designation of program codes and implements the function of the above embodiments.

[0046]

[Effects of the Invention]

Thus, in accordance with the present invention, as described above, it is possible to provide a speech synthesizing apparatus capable of selecting better phoneme units, as a result of which synthesized speech of superior quality can be produced, a method of controlling this apparatus, and a storage medium storing a program for implementing this control method.

[0047]

[Brief Description of the Drawings]

Fig. 1 is a diagram showing the construction of a speech synthesizing apparatus according to a first embodiment of the present invention;

Fig. 2 is a block diagram illustrating a functional arrangement, which relates to phoneme data selection processing for selecting optimum phoneme data from a set of phoneme data having the same phoneme environment and fundamental frequency, in the speech synthesizing apparatus according to the first embodiment of the present invention;

Fig. 3 is a flowchart illustrating a procedure relating to phoneme data selection processing for selecting optimum phoneme data from a set of phoneme data having the same phoneme environment and fundamental frequency, in the speech synthesizing apparatus according to the first embodiment of the present invention;

Fig. 4 is a block diagram illustrating a functional arrangement, which relates to phoneme data selection processing for selecting optimum phoneme data from a set of phoneme data having the same phoneme environment and fundamental frequency, in the speech synthesizing apparatus according to the second embodiment of the present invention; and

Fig. 5 is a flowchart illustrating a procedure relating to phoneme data selection processing for

selecting optimum phoneme data from a set of phoneme data having the same phoneme environment and fundamental frequency, in the speech synthesizing apparatus according to the second embodiment of the present invention.

[Description of Symbols]

101...control memory (ROM), 102...central processing unit, 103...memory (RAM), 104...disk device, 105...bus, 200...database, 201...phoneme retrieval unit, 202...storage area for retrieved-result, 203...power-penalty assignment processing unit, 204...storage area for results of assignment of penalties, 205...phoneme duration-penalty assignment processing unit, 206...sorting processing unit, 207...storage area for sorting-result, 208...representative-phoneme data determination processing unit, 209...storage area for representative phoneme data, 400...database, 401...phoneme retrieval unit, 402...storage area for retrieved-result, 403...power-penalty assignment processing unit, 404...storage area for results of assignment of penalties, 405...phoneme duration-penalty assignment processing unit, 406...sorting processing unit, 407...storage area for sorting-result, 408...representative-phoneme data determination processing unit, 409...storage area for representative phoneme data, 410...processing unit for assigning element-number penalty

ABSTRACT

[Summary]

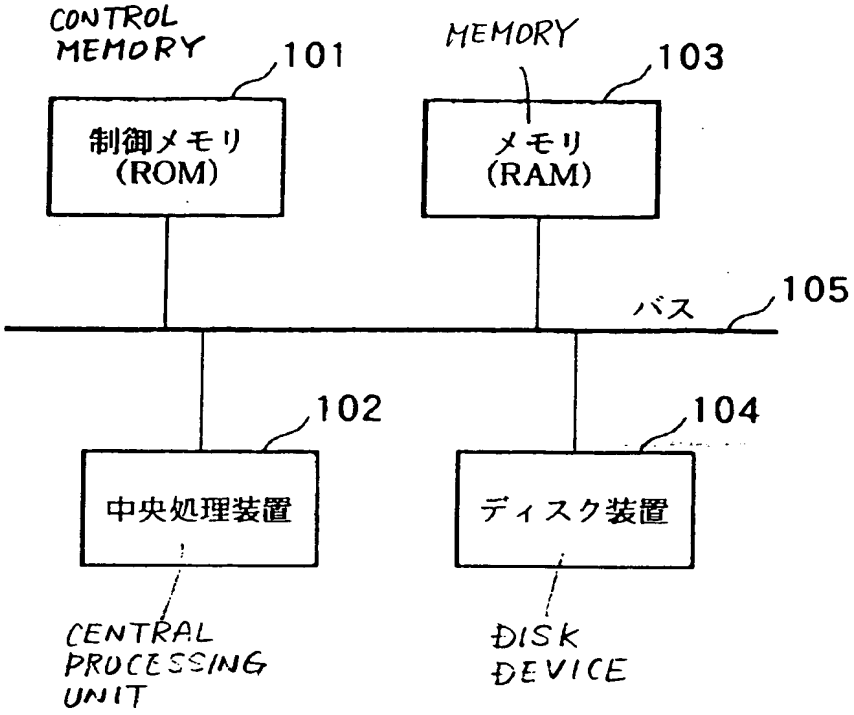
[Object] To make it possible to appropriately select phoneme data used in speech synthesis, and to suppress a decline in sound quality in speech synthesis.

[Solution] A speech synthesizing apparatus for synthesizing a speech waveform stores vocalization data, which is obtained by adding attribute information onto phoneme data, in a database 200. In accordance with prescribed retrieval conditions, a phoneme retrieval unit 201 retrieves phoneme data from the vocalization data that has been stored in the database 200 and retains the retrieved results in a retrieved-result storage area 202. A processing unit 203 for assigning a power penalty and a processing unit 205 for assigning a phoneme-duration penalty assign the penalties, on the basis of power and phoneme duration constituting the attribute information, to a set of phoneme data stored in the retrieved-result storage area. A processing unit 208 for determining typical phoneme data performs sorting on the basis of the assigned penalties and, based upon the stored results, selects phoneme data to be employed in the synthesis of a speech waveform.

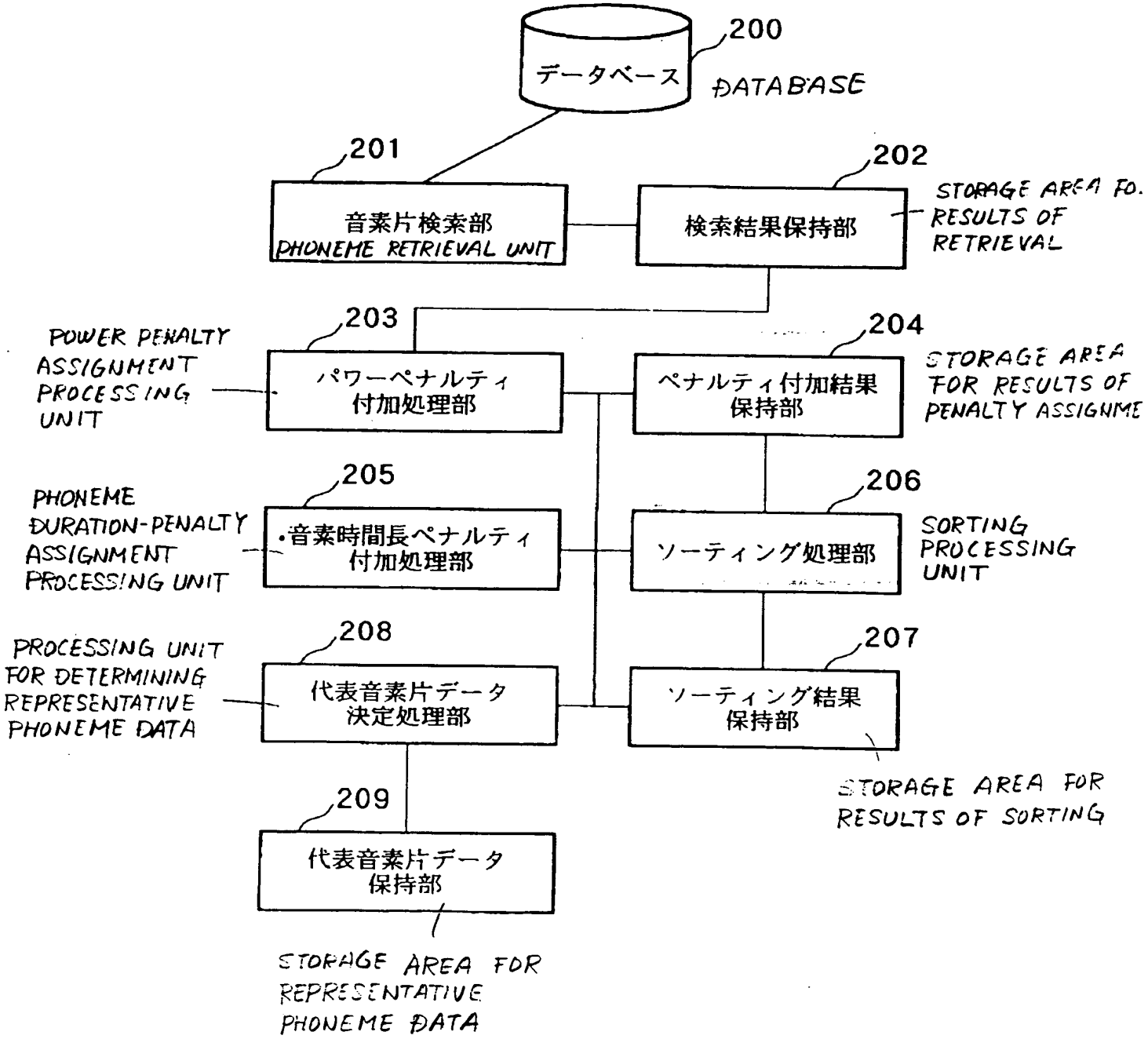
[Selected Drawing] Fig. 2.

【書類名】 図面 [Type of the Document] Drawings

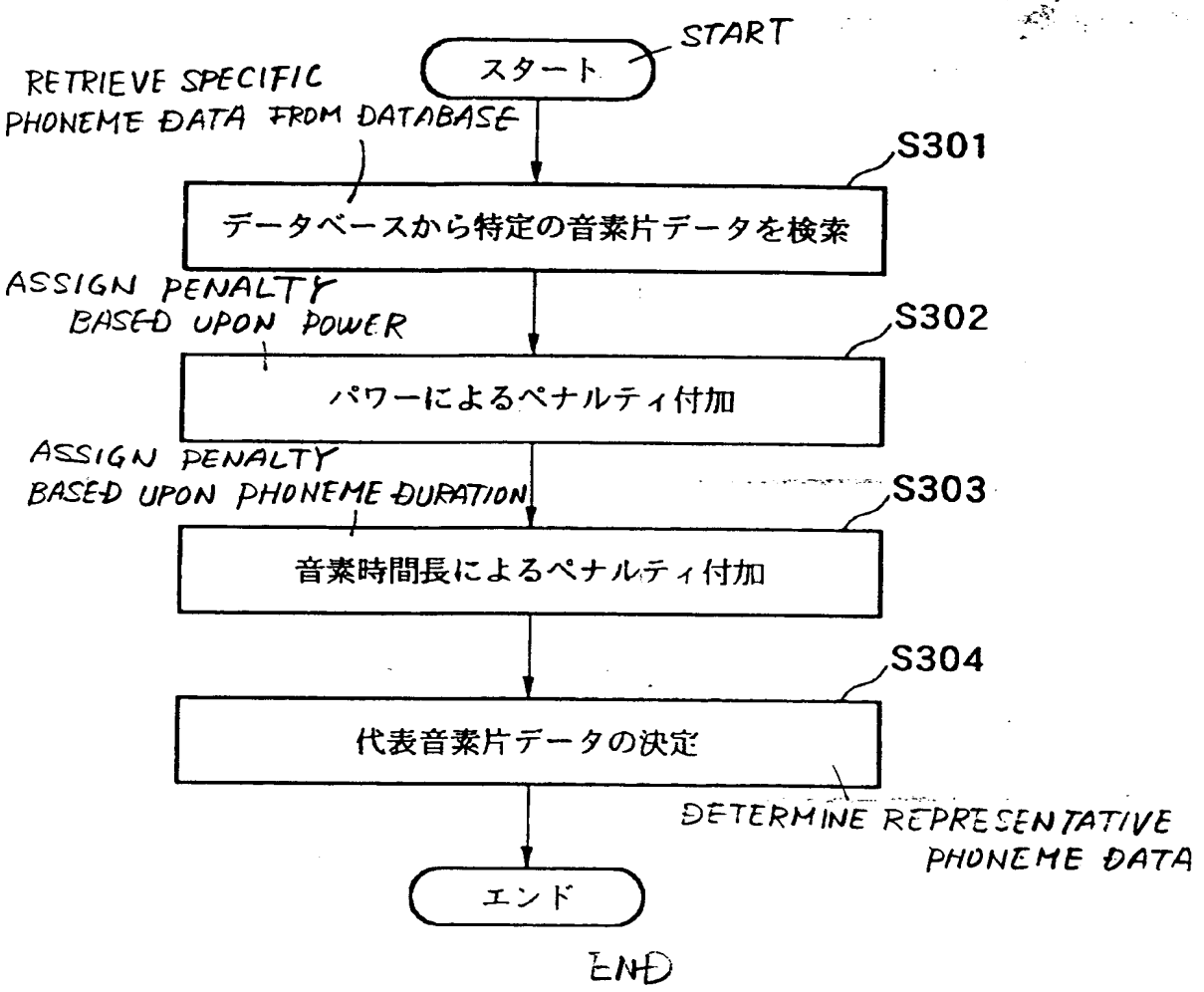
【図１】
FIG. 1



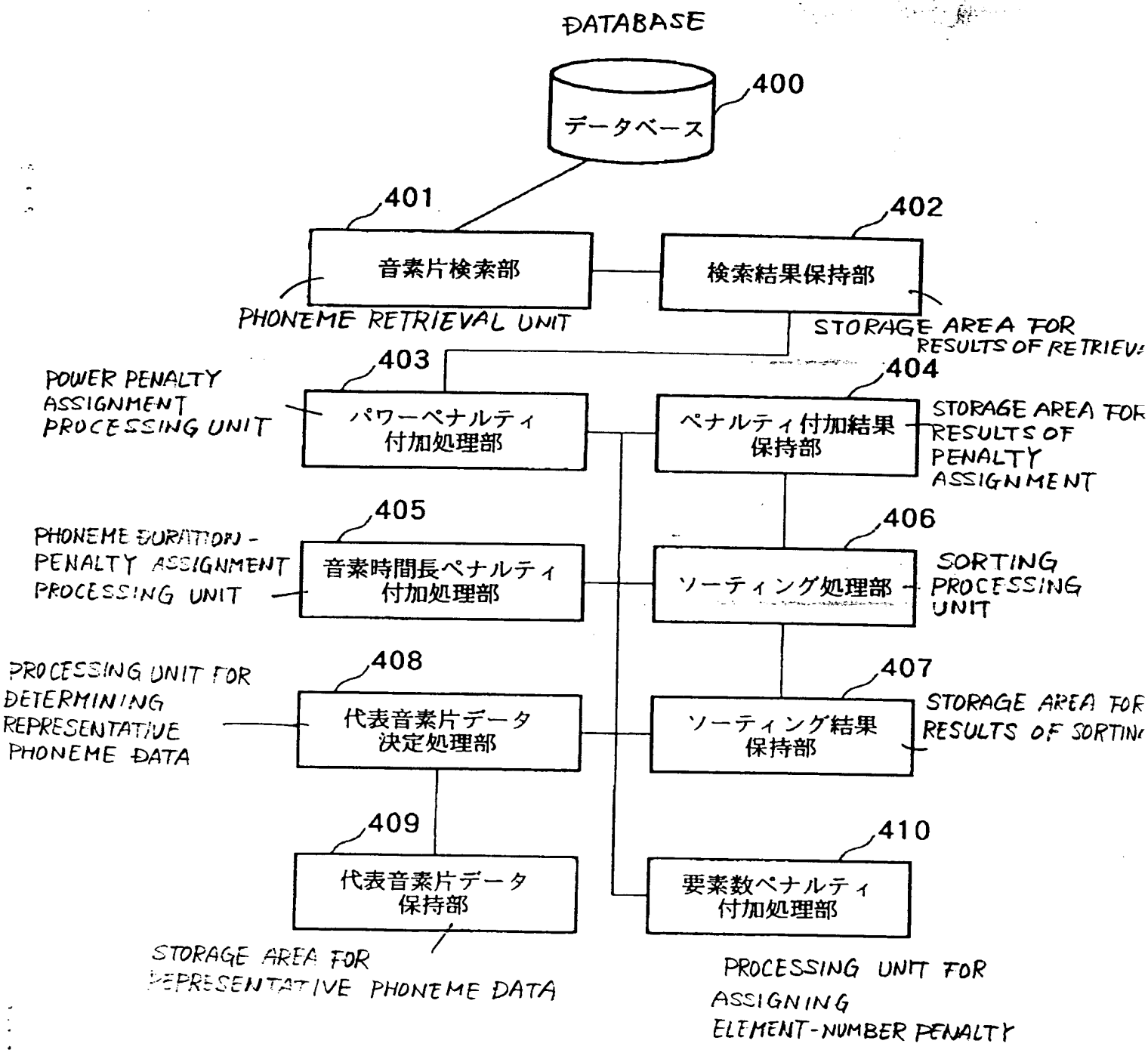
【図２】
FIG. 2


200
データベース
DATABASE
201
音素片検索部
PHONEME RETRIEVAL UNIT
202
検索結果保持部
STORAGE AREA FO.
RESULTS OF
RETRIEVAL
203
POWER PENALTY
ASSIGNMENT
PROCESSING
UNIT
パワーペナルティ
付加処理部
204
ペナルティ付加結果
保持部
STORAGE AREA
FOR RESULTS OF
PENALTY ASSIGNME
205
PHONEME
DURATION-PENALTY
ASSIGNMENT
PROCESSING UNIT
・音素時間長ペナルティ
付加処理部
206
ソーティング処理部
SORTING
PROCESSING
UNIT
208
PROCESSING UNIT
FOR DETERMINING
REPRESENTATIVE
PHONEME DATA
代表音素片データ
決定処理部
207
ソーティング結果
保持部
STORAGE AREA FOR
RESULTS OF SORTING
209
代表音素片データ
保持部
STORAGE AREA FOR
REPRESENTATIVE
PHONEME DATA

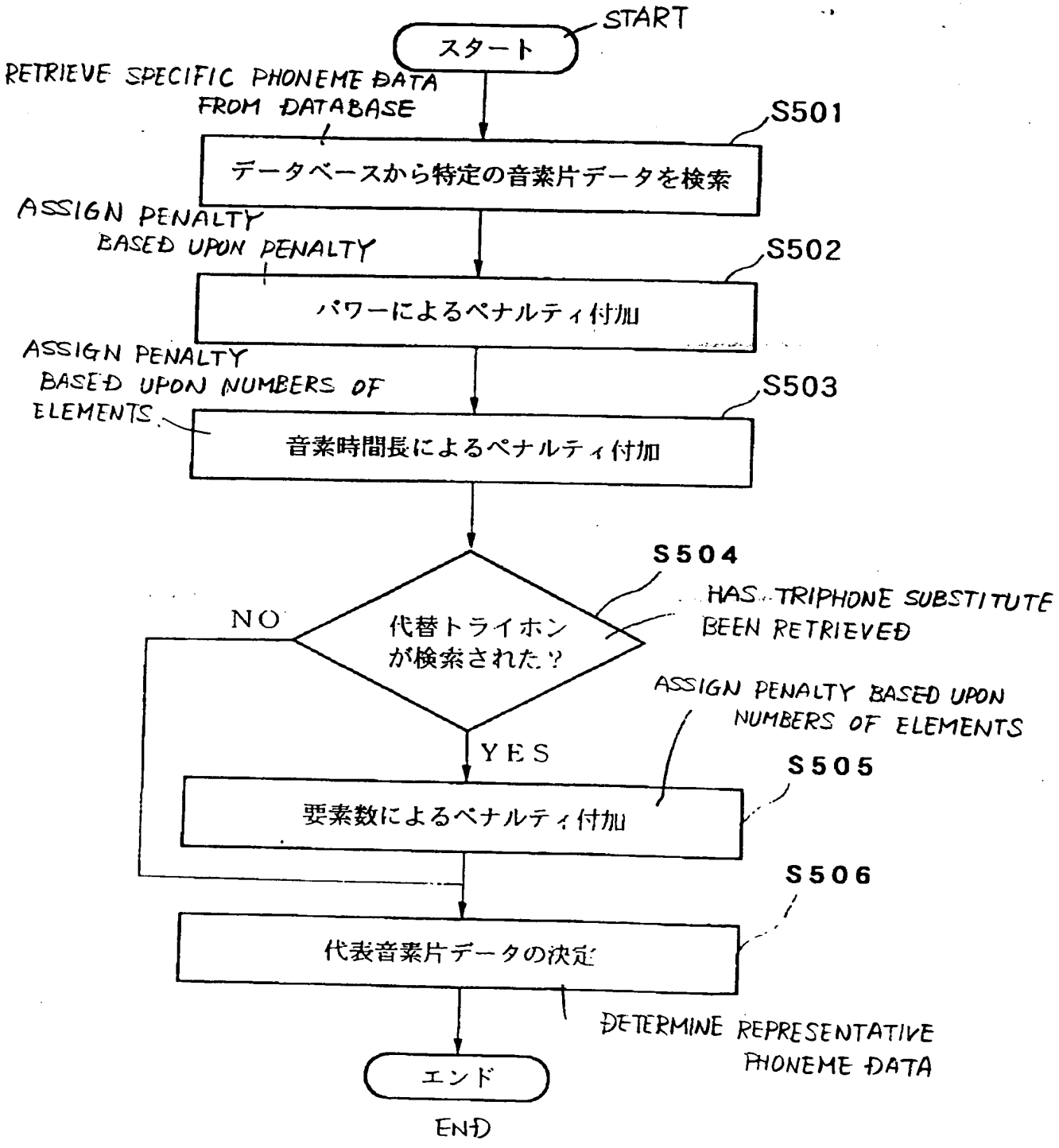
【図３】
FIG. 3


START
スタート
RETRIEVE SPECIFIC PHONEME DATA FROM DATABASE
S301
データベースから特定の音素片データを検索
ASSIGN PENALTY BASED UPON POWER
S302
パワーによるペナルティ付加
ASSIGN PENALTY BASED UPON PHONEME DURATION
S303
音素時間長によるペナルティ付加
S304
代表音素片データの決定
DETERMINE REPRESENTATIVE PHONEME DATA
エンド
END

【図４】
FIG. 4


DATABASE
400
データベース
401
音素片検索部
PHONEME RETRIEVAL UNIT
402
検索結果保持部
STORAGE AREA FOR RESULTS OF RETRIEV
403
POWER PENALTY ASSIGNMENT PROCESSING UNIT
パワーペナルティ付加処理部
404
ペナルティ付加結果保持部
STORAGE AREA FOR RESULTS OF PENALTY ASSIGNMENT
405
PHONEME DURATION-PENALTY ASSIGNMENT PROCESSING UNIT
音素時間長ペナルティ付加処理部
406
ソーティング処理部
SORTING PROCESSING UNIT
408
PROCESSING UNIT FOR DETERMINING REPRESENTATIVE PHONEME DATA
代表音素片データ決定処理部
407
ソーティング結果保持部
STORAGE AREA FOR RESULTS OF SORTIN
409
代表音素片データ保持部
STORAGE AREA FOR REPRESENTATIVE PHONEME DATA
410
要素数ペナルティ付加処理部
PROCESSING UNIT FOR ASSIGNING ELEMENT-NUMBER PENALTY

【図５】
FIG. 5


START
スタート
RETRIEVE SPECIFIC PHONEME DATA FROM DATABASE
S501
データベースから特定の音素片データを検索
ASSIGN PENALTY BASED UPON PENALTY
S502
パワーによるペナルティ付加
ASSIGN PENALTY BASED UPON NUMBERS OF ELEMENTS
S503
音素時間長によるペナルティ付加
S504
HAS TRIPHONE SUBSTITUTE BEEN RETRIEVED
代替トライホンが検索された？
NO
YES
ASSIGN PENALTY BASED UPON NUMBERS OF ELEMENTS
S505
要素数によるペナルティ付加
S506
代表音素片データの決定
DETERMINE REPRESENTATIVE PHONEME DATA
エンド
END